

## BIROn - Birkbeck Institutional Research Online

Davey, N.E. and Babu, M.M. and Blackledge, M. and Bridge, A. and Capella-Gutierrez, S. and Dosztanyi, Z. and Drysdale, R. and Edwards, R.J. and Elofsson, A. and Felli, I.C. and Gibson, T.J. and Gutmanas, A. and Hancock, J.M. and Harrow, J. and Higgins, D. and Jeffries, C.M. and Le Mercier, P. and Mészáros, B. and Necci, M. and Notredame, C. and Orchard, S. and Ouzounis, C.A. and Pancsa, R. and Papaleo, E. and Pierattelli, R. and Piovesan, D. and Promponas, V.J. and Ruch, P. and Rustici, G. and Romero, P. and Sarntivijai, S. and Saunders, G. and Schuler, B. and Sharan, M. and Shields, D.C. and Sussman, J.L. and Tedds, J.A. and Tompa, P. and Turewicz, M. and Vondrasek, J. and Vranken, W.F. and Wallace, Bonnie A. and Wichapong, K. and Tosatto, S.C.E. (2019) An intrinsically disordered proteins community for ELIXIR. F1000Research 8 , p. 1753. ISSN 2046-1402.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/30577/>

*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>  
contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).

or alternatively



# OPINION ARTICLE

## An intrinsically disordered proteins community for ELIXIR

[version 1; peer review: 2 approved]

Norman E. Davey<sup>1</sup>, M. Madan Babu<sup>2</sup>, Martin Blackledge<sup>3</sup>, Alan Bridge<sup>4</sup>, Salvador Capella-Gutierrez <sup>5</sup>, Zsuzsanna Dosztanyi<sup>6</sup>, Rachel Drysdale<sup>7</sup>, Richard J. Edwards<sup>8</sup>, Arne Elofsson <sup>9</sup>, Isabella C. Felli<sup>10</sup>, Toby J. Gibson<sup>11</sup>, Aleksandras Gutmanas <sup>12</sup>, John M. Hancock <sup>7</sup>, Jen Harrow<sup>7</sup>, Desmond Higgins<sup>13</sup>, Cy M. Jeffries<sup>14</sup>, Philippe Le Mercier <sup>4</sup>, Balint Mészáros<sup>6</sup>, Marco Necci<sup>15</sup>, Cedric Notredame<sup>16,17</sup>, Sandra Orchard <sup>18</sup>, Christos A. Ouzounis<sup>19</sup>, Rita Pancsa<sup>20</sup>, Elena Papaleo<sup>21</sup>, Roberta Pierattelli <sup>10</sup>, Damiano Piovesan<sup>15</sup>, Vasilis J. Promponas <sup>22</sup>, Patrick Ruch<sup>23</sup>, Gabriella Rustici<sup>24</sup>, Pedro Romero <sup>25</sup>, Sirarat Sarntivijai<sup>7</sup>, Gary Saunders<sup>7</sup>, Benjamin Schuler<sup>26</sup>, Malvika Sharan<sup>11</sup>, Denis C. Shields<sup>13</sup>, Joel L. Sussman<sup>27</sup>, Jonathan A. Tedds<sup>7</sup>, Peter Tompa<sup>28</sup>, Michael Turewicz <sup>29</sup>, Jiri Vondrasek<sup>30</sup>, Wim F. Vranken<sup>31</sup>, Bonnie Ann Wallace<sup>32</sup>, Kanin Wichapong<sup>33</sup>, Silvio C. E. Tosatto<sup>15</sup>

<sup>1</sup>Division of Cancer Biology, Institute of Cancer Research, UK, London, SW3 6JB, UK

<sup>2</sup>MRC Laboratory of Molecular Biology,, Cambridge, CB2 0QH, UK

<sup>3</sup>Institut de Biologie Structurale, Université Grenoble Alpes, Grenoble, 38000, France

<sup>4</sup>Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, Geneva, Switzerland

<sup>5</sup>Life Sciences Department, Barcelona Supercomputing Center (BSC), Barcelona, Spain

<sup>6</sup>Department of Biochemistry, Eötvös Loránd University, Budapest, H-1117, Hungary

<sup>7</sup>ELIXIR Hub, Wellcome Genome Campus, Cambridge, CB10 1SD, UK

<sup>8</sup>School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

<sup>9</sup>Department of Biochemistry and Biophysics and Science for Life Laboratory, Stockholm University, Stockholm, Sweden

<sup>10</sup>Department of Chemistry and CERM "Ugo Schiff", University of Florence, Florence, Italy

<sup>11</sup>Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

<sup>12</sup>Protein Data Bank in Europe, European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Cambridge, CB10 1SD, UK

<sup>13</sup>Conway Institute of Biomolecular & Biomedical Research, University College Dublin, Belfield, Dublin, D4, Ireland

<sup>14</sup>European Molecular Biology Laboratory, Hamburg, Germany

<sup>15</sup>Department of Biomedical Sciences, University of Padua, Padua, Italy

<sup>16</sup>Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, 08003, Spain

<sup>17</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain

<sup>18</sup>European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Cambridge, CB10 1SD, UK

<sup>19</sup>BCPL-CPERI, Centre for Research & Technology Hellas (CERTH), Thessalonica, 57001, Greece

<sup>20</sup>Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, Budapest, H-1117, Hungary

<sup>21</sup>Computational Biology Laboratory, Danish Cancer Society Research Center, Copenhagen, 2100, Denmark

<sup>22</sup>Bioinformatics Research Laboratory, Department of Biological Sciences, University of Cyprus, Nicosia, CY-1678, Cyprus

<sup>23</sup>HES-SO/HEG and SIB Text Mining, Swiss Institute of Bioinformatics, Geneva, Switzerland

<sup>24</sup>Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK

<sup>25</sup>University of Wisconsin-Madison, Madison, WI, 53706-1544, USA

<sup>26</sup>Department of Biochemistry, University of Zurich, Zurich, Switzerland

<sup>27</sup>Department of Structural Biology and the Israel Structural Proteomics, Center (ISPC), Weizmann Institute of Science, Rehovot, 7610001, Israel

<sup>28</sup>VIB Center for Structural Biology (CSB), VIB Flemish Institute for Biotechnology, Brussels, 1050, Belgium

<sup>29</sup>Faculty of Medicine, Medizinisches Proteom-Center, Ruhr University Bochum, GesundheitsCampus 4, Bochum, 44801, Germany

<sup>30</sup>Institute of Organic Chemistry and Biochemistry, CAS, Prague, Czech Republic

<sup>31</sup>VUB/ULB Interuniversity Institute of Bioinformatics in Brussels and Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, B-1050, Belgium

<sup>32</sup>Institute of Structural and Molecular Biology, Birkbeck College, University of London, London, WC1H 0HA, UK

<sup>33</sup>Department of Biochemistry, Cardiovascular Research Institute Maastricht (CARIM), Maastricht University, Maastricht, The Netherlands

**V1** First published: 15 Oct 2019, 8(ELIXIR):1753 (<https://doi.org/10.12688/f1000research.20136.1>)

Latest published: 15 Oct 2019, 8(ELIXIR):1753 (<https://doi.org/10.12688/f1000research.20136.1>)

## Abstract




Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are now recognised as major determinants in cellular regulation. This white paper presents a roadmap for future e-infrastructure developments in the field of IDP research within the ELIXIR framework. The goal of these developments is to drive the creation of high-quality tools and resources to support the identification, analysis and functional characterisation of IDPs. The roadmap is the result of a workshop titled “An intrinsically disordered protein user community proposal for ELIXIR” held at the University of Padua. The workshop, and further consultation with the members of the wider IDP community, identified the key priority areas for the roadmap including the development of standards for data annotation, storage and dissemination; integration of IDP data into the ELIXIR Core Data Resources; and the creation of benchmarking criteria for IDP-related software. Here, we discuss these areas of priority, how they can be implemented in cooperation with the ELIXIR platforms, and their connections to existing ELIXIR Communities and international consortia. The article provides a preliminary blueprint for an IDP Community in ELIXIR and is an appeal to identify and involve new stakeholders.

## Keywords

ELIXIR, intrinsically disordered proteins, protein-protein interactions, protein function, databases, community standards, protein dynamics, cellular regulation

## Open Peer Review

Reviewer Status  

	Invited Reviewers	
	1	2
<b>version 1</b>		
published 15 Oct 2019	report	report
<hr/>		
1	<b>Iva Pritisanac</b>  , Hospital for Sick Children, Toronto, Canada University of Toronto, Toronto, Canada <b>Julie Forman-Kay</b> , Hospital for Sick Children, Toronto, Canada University of Toronto, ON, Canada	
2	<b>Osman Ugur Sezerman</b> , Acibadem University, Istanbul, Turkey	
Any reports and responses or comments on the article can be found at the end of the article.		



This article is included in the **ELIXIR** gateway.

**Corresponding authors:** Norman E. Davey ([norman.davey@icr.ac.uk](mailto:norman.davey@icr.ac.uk)), Silvio C. E. Tosatto ([silvio.tosatto@unipd.it](mailto:silvio.tosatto@unipd.it))

**Author roles:** Davey NE: Conceptualization, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Babu MM: Conceptualization, Writing – Review & Editing; Blackledge M: Conceptualization, Writing – Review & Editing; Bridge A: Conceptualization, Writing – Review & Editing; Capella-Gutierrez S: Conceptualization, Writing – Review & Editing; Dosztanyi Z: Conceptualization, Writing – Review & Editing; Drysdale R: Conceptualization, Writing – Review & Editing; Edwards RJ: Conceptualization, Writing – Review & Editing; Elofsson A: Conceptualization, Writing – Review & Editing; Felli IC: Conceptualization, Writing – Review & Editing; Gibson TJ: Conceptualization, Writing – Review & Editing; Gutmanas A: Conceptualization, Writing – Review & Editing; Hancock JM: Conceptualization, Writing – Review & Editing; Harrow J: Conceptualization, Writing – Review & Editing; Higgins D: Conceptualization, Writing – Review & Editing; Jeffries CM: Conceptualization, Writing – Review & Editing; Le Mercier P: Conceptualization, Writing – Review & Editing; Mészáros B: Conceptualization, Writing – Review & Editing; Necci M: Conceptualization, Writing – Review & Editing; Notredame C: Conceptualization, Writing – Review & Editing; Orchard S: Conceptualization, Writing – Review & Editing; Ouzounis CA: Conceptualization, Writing – Review & Editing; Pancsa R: Conceptualization, Writing – Review & Editing; Papaleo E: Conceptualization, Writing – Review & Editing; Pierattelli R: Conceptualization, Writing – Review & Editing; Piovesan D: Conceptualization, Writing – Review & Editing; Promponas VJ: Conceptualization, Writing – Review & Editing; Ruch P: Conceptualization, Writing – Review & Editing; Rustici G: Conceptualization, Writing – Review & Editing; Romero P: Conceptualization, Writing – Review & Editing; Sarntivijai S: Conceptualization, Writing – Review & Editing; Saunders G: Conceptualization, Writing – Review & Editing; Schuler B: Conceptualization, Writing – Review & Editing; Sharan M: Conceptualization, Writing – Review & Editing; Shields DC: Conceptualization, Writing – Review & Editing; Sussman JL: Conceptualization, Writing – Review & Editing; Tedds JA: Conceptualization, Writing – Review & Editing; Tompa P: Conceptualization, Writing – Review & Editing; Turewicz M: Conceptualization, Writing – Review & Editing; Vondrasek J: Conceptualization, Writing – Review & Editing; Vranken WF: Conceptualization, Writing – Review & Editing; Wallace BA: Conceptualization, Writing – Review & Editing; Wichapong K: Conceptualization, Writing – Review & Editing; Tosatto SCE: Conceptualization, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** A strategic meeting to identify the key priority areas for the roadmap was funded by ELIXIR, the research infrastructure for life-science data. BW was funded by a grant from the BBSRC [BB/P024092/1]. AB, SO and PLM were supported by the National Eye Institute (NEI), National Human Genome Research Institute (NHGRI), National Heart, Lung, and Blood Institute (NHLBI), National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of General Medical Sciences (NIGMS), and National Institute of Mental Health (NIMH) of the National Institutes of Health under Award Number [U24HG007822] (the content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health), by European Molecular Biology Laboratory (EMBL) and by the Swiss Federal Government through the State Secretariat for Education, Research and Innovation SERI. AG was supported by the European Molecular Biology Laboratory (EMBL) and the Wellcome Trust [104948]. AE was supported by a Vetenskapsrådet grant [2016-03798]. EP was supported by Danmarks Grundforskningsfond [DNRF125] and a Carlsberg Foundation Distinguished Fellowship [CF18-0314]. JLS was supported by the Israel I-CORE Project: Integrated Structural Cell Biology and by Instruct-ERIC. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement [778247] and COST Action BM1405 NGP-net.

**Copyright:** © 2019 Davey NE *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Davey NE, Babu MM, Blackledge M *et al.* **An intrinsically disordered proteins community for ELIXIR [version 1; peer review: 2 approved]** F1000Research 2019, 8(ELIXIR):1753 (<https://doi.org/10.12688/f1000research.20136.1>)

**First published:** 15 Oct 2019, 8(ELIXIR):1753 (<https://doi.org/10.12688/f1000research.20136.1>)

## Introduction

Intrinsically disordered regions (IDRs), protein segments that lack persistent secondary or tertiary structure (Chouard, 2011; Dyson & Wright, 2005; Tompa, 2011), are predicted to cover almost a third of the residues in eukaryotic proteomes (Pancsa & Tompa, 2012; Xue *et al.*, 2012). IDRs play a central role in cell regulation and contribute significantly to the cellular complexity of higher eukaryotes (Dunker *et al.*, 2008; Dyson & Wright, 2005; Forman-Kay & Mittag, 2013; Gouw *et al.*, 2018; Mitrea & Kriwacki, 2016; Schad *et al.*, 2018; Tompa, 2005; Van Roey *et al.*, 2014). They represent a major source of protein diversity and versatility on the level of organisms and during evolution (Babu *et al.*, 2012; Buljan *et al.*, 2012; Davey *et al.*, 2015; Light *et al.*, 2013; Weatheritt & Gibson, 2012). In the human proteome, IDRs are expected to contain up to one hundred thousand interaction interfaces and a million sites of post-translational modification (Tompa *et al.*, 2014). However, to date, only a small fraction of these functional modules have been characterised (Gouw *et al.*, 2018; Schad *et al.*, 2018).

IDR-mediated interactions are commonly found in dynamic and transient complexes that underlie enzyme inhibition, signal transduction and liquid-liquid phase transition (Borgia *et al.*, 2018; Ivarsson & Jemth, 2019; Kriwacki *et al.*, 1996; Martin & Mittag, 2018; Mitrea & Kriwacki, 2016; Olsen *et al.*, 2017; Scott & Pawson, 2009; Wright & Dyson, 2015). IDR-mediated interactions are also major determinants of protein regulation, and are so far known to affect: (i) the post-translational modification state of a protein by acting as docking sites to recruit modifying enzymes; (ii) the cellular half-life of a protein by recruiting E3 ubiquitin ligases resulting in ubiquitin-dependent proteasomal degradation of the protein; and (iii) the localisation of a protein by acting as signals that target proteins to specific subcellular locations (Beltrao *et al.*, 2012; Davey & Morgan, 2016; Gouw *et al.*, 2018; Guharoy *et al.*, 2016; Iakoucheva *et al.*, 2004; Mészáros *et al.*, 2017; Van Roey *et al.*, 2014). Many functions of IDRs are directly associated with their structural attributes, without directly contributing to binding events that result in complex formation; for example, IDRs can act as entropic springs, flexible linkers or spacers (Tompa, 2005; van der Lee *et al.*, 2014), or fly-casting regions to capture binding partners (Shoemaker *et al.*, 2000). Furthermore, IDRs are subject to extensive pre- and post-translational regulation to modulate protein function in response to cellular stimuli (Bah & Forman-Kay, 2016; Csizsmok & Forman-Kay, 2018; Van Roey *et al.*, 2013; Van Roey *et al.*, 2012; Weatheritt *et al.*, 2012).

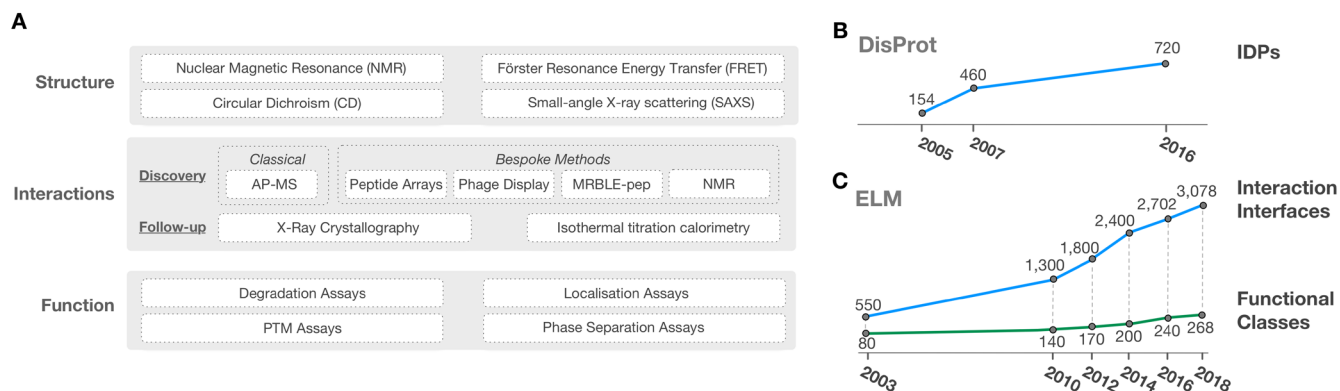
As a result of the fundamental regulatory functions performed by IDRs, and the cell-state conditionality of these regulatory processes, IDRs encode many of critical steps of a protein life-cycle from the ribosome to the proteasome. Consequently, IDRs play a key role in many human diseases, including cancer (p53), Alzheimer's disease (A $\beta$ , Tau) or Parkinson's disease ( $\alpha$ -synuclein) (Shigemitsu & Hiroaki, 2018; Uversky *et al.*, 2008) and human IDR interfaces are often mimicked by pathogens to hijack host pathways and deregulate the cell (Davey *et al.*, 2011; Dyson & Wright, 2018; Kruse *et al.*, 2019;

Via *et al.*, 2015; Xue *et al.*, 2012). Given their therapeutic relevance, IDP-mediated interactions are now seen as potential drug targets (Corbi-Verge & Kim, 2016). Therefore, a better understanding of their structure and function will help to develop new strategies to fight human diseases.

IDP research spans several experimental fields studying protein structure and function including structural biology, biophysics, biochemistry, cell biology, proteomics, comparative genomics, systems biology, synthetic biology and pharmacology (Figure 1A) (Blikstad & Ivarsson, 2015; Corbi-Verge & Kim, 2016; Felli & Pierattelli, 2015; Forman-Kay & Mittag, 2013; Plitzko *et al.*, 2017). On a structural level, IDRs do not adopt a single stable highly populated structure, instead, they are structurally heterogeneous, continuously sample a wide ensemble of conformations, with preferentially sampled intramolecular contacts driving local transient secondary structure and compaction of conformations (Davey, 2019; Dyson & Wright, 2005; Forman-Kay & Mittag, 2013; Holehouse & Pappu, 2018). Therefore, IDRs are more easily described by probabilistic models than the intuitive visual representations of structures of folded protein regions. Classical methods for structural characterisation of a protein, such as X-Ray crystallography, are unable to capture the dynamic structures of IDRs. Instead, the structural aspects of the IDRs are studied by a range of biophysical methods including Nuclear Magnetic Resonance (NMR), Small-angle X-ray scattering (SAXS), circular dichroism (CD) or Förster resonance energy transfer (FRET) (Felli & Pierattelli, 2015; Fuertes *et al.*, 2017; Holmstrom *et al.*, 2018; Plitzko *et al.*, 2017; Tompa, 2011; Tolchard *et al.*, 2018). The results of these complementary methods can be integrated to build a model of IDR structure and dynamics.

The functional aspects of IDPs are studied by a battery of well-established cell biology and biophysical approaches (Gibson *et al.*, 2015) (Figure 1). Many of the approaches were developed for protein interaction elucidation, for example, by coupling mutagenesis to affinity-purification. However, interactions mediated by IDRs are often defined by their low affinity and cell-state dependent conditionality, two properties that are not always amenable to the available experimental protein interaction detection methods (Gibson *et al.*, 2015; Van Roey *et al.*, 2014). Therefore, there are a growing number of approaches that have been specifically designed to characterise low affinity IDR-mediated interactions such as peptide arrays, proteomic phage display (ProP-PD) and peptides attached to Microspheres with Ratiometric Barcode Lanthanide Encoding (MRBLE-pep) (Blikstad & Ivarsson, 2015; Davey *et al.*, 2017; Nguyen *et al.*, 2017; Volkmer, 2009). As IDRs often adopt a dominant conformation in their bound state, many IDR-containing interfaces can be studied in complex with the structured binding partners using structural approaches including NMR spectroscopy and X-Ray Crystallography (Bonetti *et al.*, 2018; Fuxreiter, 2018; Iešmantavicius *et al.*, 2014; Schad *et al.*, 2018). Distinct IDR functionalities are also studied using isolation, deletion or mutagenesis of a functional module combined with bespoke





**Figure 1. (A)** Key experimental methods used across the three major research focuses of the IDP field: structure, interactome and function. **(B)** The growth in structural IDP data curated by the DisProt database (Piovesan *et al.*, 2017). DisProt was established over a decade ago in the USA, and recently brought to Europe after years of inactivity and completely re-annotated, this explains the lag in the curation. A huge amount of IDP/IDR data in the IDP literature remain uncured and the vast majority of IDP regions remain to be characterised (Pancsa & Tompa, 2012; Xue *et al.*, 2012). **(C)** The growth in functional IDP data curated by the Eukaryotic Linear Motif (ELM) resource (Gouw *et al.*, 2018). Similar to the structural IDR data, a huge amount of functional IDP/IDR data in the IDP literature remains uncured and the vast majority of functional modules in IDPs remain to be characterised (Tompa *et al.*, 2014).

assays to characterise phase separation, protein localisation, stability and post-translational modification (Gibson *et al.*, 2015; Gouw *et al.*, 2018; Martin & Mittag, 2018).

The computational IDP field tackles several distinct research tasks including: (i) the prediction of IDRs (Cilia *et al.*, 2014; Dosztányi *et al.*, 2005), (ii) the identification of functional modules in IDRs (Dosztányi *et al.*, 2009; Edwards & Palopoli, 2015; Krystkowiak & Davey, 2017; Neduva & Russell, 2006), (iii) docking of peptides that are intrinsically disordered in their unbound state (Raveh *et al.*, 2011; Trabuco *et al.*, 2012), (iv) force field development and molecular simulations for IDR structure (Best, 2017; Chong *et al.*, 2017; Huang & MacKerell, 2018; Stanley *et al.*, 2015), (v) *in silico* inhibitor design and development for IDRs (Baggett & Nath, 2018; Santofimia-Castaño *et al.*, 2019; Yu *et al.*, 2016) and (vi) the processing of the data produced by experimental structural and functional analyses of IDRs (Bernadó *et al.*, 2007; Franke *et al.*, 2017; Nodet *et al.*, 2009; Ozenne *et al.*, 2012) (Table 1). The members of the European IDP community are major contributors across these computational IDP fields (Figure 2).

Europe hosts many of the key IDP resources (Table 2). DisProt is the largest database of manually curated experimental data on IDP structure (Piovesan *et al.*, 2017; Sickmeier *et al.*, 2007). MobiDB and D<sup>2</sup>P<sup>2</sup> are central resources of pre-computed structural prediction of IDPs (Piovesan *et al.*, 2018; Oates *et al.*, 2013). ELM is a manually curated database of binding regions residing in IDPs (Gouw *et al.*, 2018). SASBDB and PCDDDB are central resources for Small-angle X-ray Scattering (SAXS) and Circular Dichroism (CD) data (Valentini *et al.*, 2015; Whitmore *et al.*, 2017). PED<sup>3</sup> is a database of conformational ensembles from Nuclear Magnetic Resonance (NMR) and SAXS data, and Molecular Dynamics (MD) simulations (Varadi *et al.*, 2014). It is important to note that the vast majority

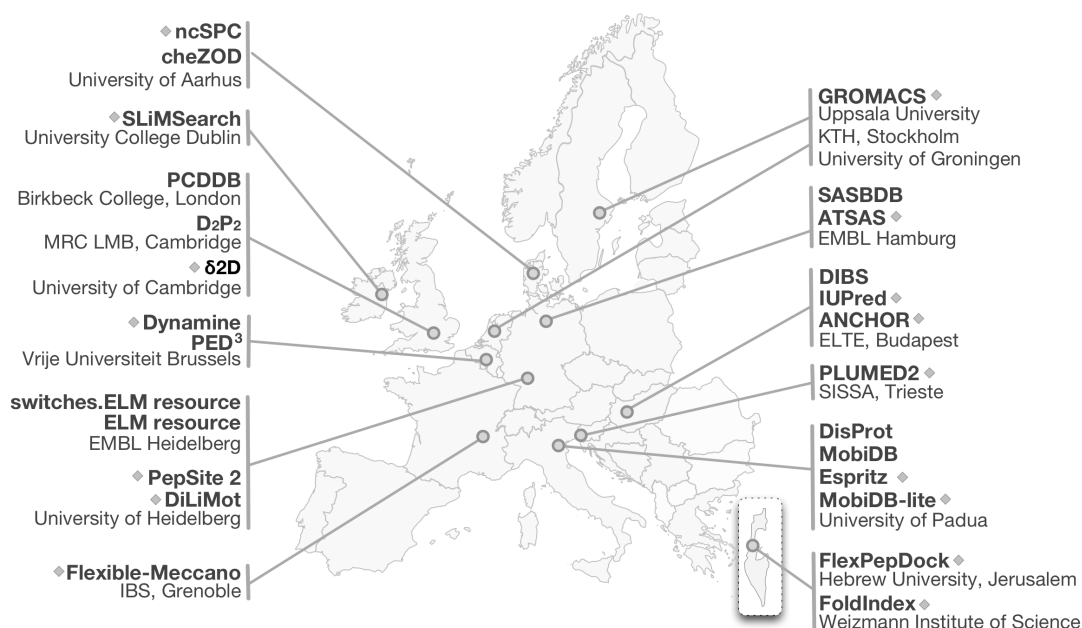
of the computational frameworks developed by the IDP field had to be created from scratch given the unique structural and functional aspects of IDRs. For example, the analysis of IDP structure drew predominantly from polymer biophysics rather than the existing structural biology of folded proteins (Holehouse & Pappu, 2018; Milles *et al.*, 2018; Schuler *et al.*, 2016).

The ELIXIR IDP Community has grown out of the NGP-NET COST Action (Non-globular proteins: From sequence to structure, function and application in molecular physiopathology), a scientific cooperation funded in 2015 under a Horizon 2020 EU Framework Programme. The NGP-NET community spanned 30 different countries, plus EMBL-EBI and EMBL Heidelberg. NGP-NET held a series of thematic workshops on IDPs to drive the development of computational resources and community standards. A strategic workshop titled “Intrinsically Disordered Proteins in Core Data Resources” was organised by NGP-NET at the EBI campus, Hinxton, the UK on June 1–2, 2017 to discuss the integration of IDP-related data and computational resources into the ELIXIR framework. A major outcome of this workshop was the recognition that IDP annotations are significantly underrepresented in the ELIXIR Core Data Resources (CDRs) (Durinx *et al.*, 2016). These initial insights evolved into a comprehensive plan to develop the IDP field and integrate key IDP resources and tools into the CDRs. That plan formed the basis for the ELIXIR IDP Community proposal that was submitted and presented at the ELIXIR Head of Nodes meeting in Basel, Switzerland on September 13th, 2017. After completing the ELIXIR Community application process, including the creation of this white paper, the IDP Community proposal was accepted by the ELIXIR Head of Nodes on May 14th, 2019.

In parallel with these developments, the IDP research community has developed closer integration with ELIXIR activities.

**Table 1. Representative core software of the IDP field developed and hosted in Europe.**

Name	Description
IUPred	Prediction of intrinsically disordered regions. URL: <a href="https://iupred2a.elte.hu">https://iupred2a.elte.hu</a> (Dosztányi <i>et al.</i> , 2005)
ESpritz	Prediction of intrinsically disordered regions URL: <a href="http://protein.bio.unipd.it/espritz/">http://protein.bio.unipd.it/espritz/</a> (Walsh <i>et al.</i> , 2012)
FoldIndex	Prediction of intrinsically disordered regions URL: <a href="https://fold.weizmann.ac.il/fldbin/findex">https://fold.weizmann.ac.il/fldbin/findex</a> (Prilusky <i>et al.</i> , 2005)
MobiDB-lite	Prediction of intrinsically disordered regions URL: <a href="http://protein.bio.unipd.it/mobidblite/">http://protein.bio.unipd.it/mobidblite/</a> (Necci <i>et al.</i> , 2017)
DynaMine	Prediction of protein backbone dynamics URL: <a href="https://dynamine.ibsquares.be">https://dynamine.ibsquares.be</a> (Cilia <i>et al.</i> , 2014)
DiLiMot	Prediction of functional modules in intrinsically disordered regions. URL: <a href="http://dilimot.russelllab.org/">http://dilimot.russelllab.org/</a> (Neduva & Russell, 2006)
SLIMSearch	Prediction of functional modules in intrinsically disordered regions. URL: <a href="http://slim.ucd.ie/slimsearch/">http://slim.ucd.ie/slimsearch/</a> (Krystkowiak & Davey, 2017)
ANCHOR	Prediction of functional modules in intrinsically disordered regions. URL: <a href="https://iupred2a.elte.hu">https://iupred2a.elte.hu</a> (Dosztányi <i>et al.</i> , 2009)
PepSite 2	Prediction of IDP binding sites. URL: <a href="http://pepsite2.russelllab.org/">http://pepsite2.russelllab.org/</a> (Trabuco <i>et al.</i> , 2012)
FlexPepDock	Docking of IDPs to their ordered binding partner. URL: <a href="http://flexpepdock.furmanlab.cs.huji.ac.il/">http://flexpepdock.furmanlab.cs.huji.ac.il/</a> (Raveh <i>et al.</i> , 2011)
δ2D	Secondary structure propensity from NMR data URL: <a href="http://www-mvsoftware.ch.cam.ac.uk/">http://www-mvsoftware.ch.cam.ac.uk/</a> (Camilloni <i>et al.</i> , 2012)
ncSPC	Structural propensity calculator from NMR data URL: <a href="http://linuxnmr02.chem.rug.nl/ncSPC">http://linuxnmr02.chem.rug.nl/ncSPC</a> (Tamiola & Mulder, 2012)
Flexible-Meccano	Generation of ensemble descriptions of IDPs URL: <a href="http://www.ibs.fr/research/scientific-output/software/flexible-meccano/">http://www.ibs.fr/research/scientific-output/software/flexible-meccano/</a> (Ozenne <i>et al.</i> , 2012)
ATSAS	A data analysis suite for SAXS data including the EOM tool URL: <a href="https://www.embl-hamburg.de/biosaxs/">https://www.embl-hamburg.de/biosaxs/</a> (Bernadó <i>et al.</i> , 2007; Franke <i>et al.</i> , 2017)
GROMACS	Molecular Dynamics software applicable to IDPs URL: <a href="http://www.gromacs.org/">http://www.gromacs.org/</a> (Hess <i>et al.</i> , 2008)
PLUMED2	Library for free energy calculations URL: <a href="https://www.plumed.org/">https://www.plumed.org/</a> (Tribello <i>et al.</i> , 2014)

**Figure 2. Representative core software (Table 1, marked with a diamond) and resources (Table 2) for the IDP field developed and hosted in Europe.**

**Table 2. Representative core resources of the IDP field developed and hosted in Europe.**

Name	Description
DisProt	Curated database of experimentally validated intrinsically disordered regions. URL: <a href="http://www.disprot.org">http://www.disprot.org</a> (Piovesan <i>et al.</i> , 2017)
MobiDB	Database of predicted intrinsically disordered regions. URL: <a href="http://mobidb.bio.unipd.it">http://mobidb.bio.unipd.it</a> (Piovesan <i>et al.</i> , 2018)
D2P2	Database of predicted intrinsically disordered regions. URL: <a href="http://d2p2.pro/">http://d2p2.pro/</a> (Oates <i>et al.</i> , 2013)
PED3	Database of IDP conformational ensembles. URL: <a href="http://pedb.vib.be/">http://pedb.vib.be/</a> (Varadi <i>et al.</i> , 2014)
CheZOD	Database of structural propensities from NMR chemical shift data. URL: <a href="http://www.protein-nmr.org/">http://www.protein-nmr.org/</a> (Nielsen & Mulder, 2016)
SASBDB	Repository for Small-angle X-ray Scattering (SAXS) data. URL: <a href="https://www.sasbdb.org/">https://www.sasbdb.org/</a> (Valentini <i>et al.</i> , 2015)
PCDDb	Repository for Circular Dichroism (CD) data. URL: <a href="http://pcddb.cryst.bbk.ac.uk/">http://pcddb.cryst.bbk.ac.uk/</a> (Whitmore <i>et al.</i> , 2017)
ELM	Curated database of Short, Linear Motifs. URL: <a href="http://elm.eu.org/">http://elm.eu.org/</a> (Gouw <i>et al.</i> , 2018)
DIBS	Curated database of intrinsically disordered binding regions. URL: <a href="http://dibs.enzim.ttk.mta.hu/">http://dibs.enzim.ttk.mta.hu/</a> (Schad <i>et al.</i> , 2018)
Switches.ELM	Curated conditional interactomics database for IDPs. URL: <a href="http://switches.elm.eu.org">http://switches.elm.eu.org</a> (Van Roey <i>et al.</i> , 2013)

A follow up strategic meeting, “The 2<sup>nd</sup> Workshop on Intrinsically Disordered Proteins in Core Data Resources”, was held in Prague on March 13–14, 2019 and brought together data producers, database developers and a representative of the ELIXIR Data Platform to discuss the integration of IDP data into the CDRs; and a member of the IDP research community attended the ELIXIR Interoperability Platform Face To Face Meeting on 1–2 April 2019. Furthermore, several actions have already taken place within the ELIXIR framework to tackle time-sensitive priorities including two implementation studies, “Implementation study for the integration of ELIXIR-IIB in ELIXIR Data Curation activities” and “Integration and standardisation of intrinsically disordered protein data”, focussing on interoperability. An additional, fundamental development associated with these implementation studies was the initialisation of the HUPO (Human Proteome Organisation) Proteomics Standards Initiative (PSI) Intrinsic Disorder workgroup.

The goal of the IDP Community in ELIXIR is to support the development of standards, tools and resources to accelerate the identification, analysis and functional characterisation of intrinsically disordered regions. Here, we introduce the major areas of priority for the development of an e-infrastructure that will allow the community to realise these goals while supporting the needs of the generators, users and consumers of IDP data.

### Identification of community challenges

A strategic workshop titled “An intrinsically disordered protein user community proposal for ELIXIR” was held on October 31st, 2018 at the University of Padua, Italy. The meeting was attended by participants from 13 ELIXIR nodes: Belgium, Cyprus, Czech Republic, EBI, Germany, Hungary, Ireland, Israel, Italy, Netherlands, Spain, Switzerland, and the United Kingdom. ELIXIR members in both Greece and Sweden were also interested, but could not attend. The ELIXIR Hub team was represented by John Hancock, the ELIXIR Communities and Services Coordinator and representatives from the Interoperability, Tools and Training ELIXIR platforms. Members of the 3DBioInfo and Proteomics ELIXIR Communities, the Instruct-ERIC structural biology research infrastructure, and the ELIXIR CDRs of high relevance to the IDP research, namely PDBe (Mir *et al.*, 2018), IMEx/IntAct (Orchard *et al.*, 2014) and UniProt (UniProt Consortium, 2019), were also present. During the meeting, the following topics were identified as key areas of priority for the ELIXIR IDP Community.

#### Area 1 – Standards and exchange formats for IDP data

The IDP field has not yet established official standards to allow consistent storage and dissemination of data. Hence, a key responsibility of the IDP Community is the definition of guidelines and standards to improve the reproducibility, interpretation, and dissemination of experimental data. The



development of novel standards would result in a shift from the current “organic” interoperability of the community, where each group defines their own formats and creates a range of parsers to read the formats developed by other groups, to standardised data dissemination accessible to the larger biological community. For functional data describing protein interactions, molecular interaction interchange standards based on the guidelines of the HUPO-PSI-MI molecular standards (Sivade Dumousseau *et al.*, 2018) can be applied. However, for structural data, no pre-existing solution exists. In the simplest cases, experimental evidence for IDRs can be mapped to protein sequences as sequence features. Structural descriptions such as IDR conformational ensembles (e.g. based on experimental NMR, SAXS or MD simulations data) or secondary structural propensities (e.g. based on NMR data) will require more descriptive standards, especially if they are probabilistic. Only a subset of the required experimental and functional ontologies are already available and most IDP data cannot be described by pre-existing terminology. IDPs will require novel entries to controlled vocabularies describing specific experimental methods and protein structural concepts.

The first step of this process was undertaken at a workshop for the ELIXIR implementation study: “Integration and standardisation of intrinsically disordered protein data” held at the University of Padua, October 29–30th, 2018. The meeting proposed a first draft Minimum Information About Disorder Experiments (MIADEx) standard defining the data required for reporting an IDP experiment. A further key outcome of the meeting was the establishment of the HUPO-PSI intrinsic disorder (HUPO-PSI ID) workgroup to drive the development of the required standards, storage and dissemination formats and controlled vocabularies for the community. Upon completion, the HUPO-PSI ID recommendations will be adopted by the key community resources promoting effortless integration of IDP data into the ELIXIR CDRs. An important goal of the roadmap will be the initiation of collaborations with the experimental communities for each of the distinct structural and functional methods used in IDP research to develop: (i) experimental method specific standards and (ii) workflows with minimal required experimental detail for the characterisation of IDPs. These developments would require extensive collaboration with the data generators and the experimental method specific data repositories including SASBDB (Valentini *et al.*, 2015), PCDDDB (Whitmore *et al.*, 2017) and the BMRB (Ulrich *et al.*, 2008) to allow data produced by IDP researchers to be stored and disseminated in the most descriptive and efficient way possible. This will simplify the integration of the results of these analyses into the community resources. The recent development of the PDB-Dev repository of non-atomistic or part-atomistic structural data can provide a prototype for the deposition of experimental IDP data (Vallat *et al.*, 2018; Peng *et al.*, 2019).

## Area 2 - Automated and community-driven curation

The vast majority of the experimental data describing IDPs, the functional modules encoding their function, the regulatory mechanisms conditionally controlling that function, and

their dysregulation in disease, is isolated in the text of research and review articles and in poorly formatted supplementary tables. This hampers the integration of IDP information with data such as protein function, modification, splice variants and disease-causing single-nucleotide polymorphisms (SNPs). Consequently, data created at great expense is significantly underutilised. Annotation of bona fide IDPs, especially for their function, is currently a labour-intensive process. The IDP community has already been experimenting with crowd-sourced curation, a topic which is of interest to the ELIXIR Data Platform, and included within Task 3 of the [ELIXIR Data Platform 2019–2023 ELIXIR Programme](#).

DisProt is a successful example, leveraging the collective expertise of around 40 researchers from a dozen different IDP labs in as many countries. The rate and accuracy of curation can be improved by integrating the available ELIXIR e-infrastructure into the curation process through partial annotation of articles and through automatic selection, classification and prioritisation of the relevant articles for curation. For example, the annotation of IDP-mediated interactions directly into the IMEx consortium annotation portal would provide a pre-built environment for such an endeavour. Automatic triage and pre-compiling of data based on Europe PMC data would also greatly boost productivity, allowing the community to cope with the increasing amount and complexity of data being published (Britan *et al.*, 2018). The large number of articles describing the structure and function of IDPs has highlighted the need to exploit text-mining approaches in order to eventually leverage automatic annotation from the literature. ELIXIR would facilitate integration with existing text-mining frameworks. Finally, a future goal of the community is the early capture of IDP data pre-publication directly from the data producers to reduce the need for manual curation. The coordinated experimental and computational fields within the ELIXIR IDP Community can provide a single contact point to lobby journals to require data deposition prior to publication.

## Area 3 - Integration with ELIXIR Core Data Resources

IDP annotations are significantly underrepresented in the ELIXIR CDRs and a key goal of the ELIXIR IDP community is to facilitate the integration of IDP data and services into these resources. The CDRs do not currently annotate or import experimental data describing IDR structure and function despite their high abundance and functional importance. Recently, InterPro, PDBE, and UniProt adopted IDP predictions from sequence retrieved from MobiDB, as part of an ELIXIR Implementation Study “Integration and standardisation of intrinsically disordered protein data”. This work can be used as a blueprint for the successful integration of additional IDP data into further CDRs. It would be advantageous both for the IDP Community and the CDRs to comprehensively integrate the available IDP data into these resources. Interoperability can be guaranteed by the implementation of IDP specific ontologies to describe IDP specific features and the availability of standards-compliant RESTful APIs, enabling cross-linking and programmatic access to IDP resources. Many of the key IDP resources already host RESTful APIs and utilise persistent identifiers from CDRs. However, these APIs are currently not fully

interoperable with each other without bespoke adaptors, though this will be tackled by Priority Area 1 “Standards and exchange formats for IDP data”. The development of curation guidelines and standards in line with the requirements of the CDRs, or the adoption of CDR guidelines, will streamline the integration process. The IDP resources that annotate functional modules and their interactions are currently not members of the International Molecular Exchange (IMEx) consortium (Orchard *et al.*, 2012). A key step on the roadmap is joint curation efforts between the ELIXIR IDP Community and the IMEx consortium of IDP-mediated interactions, thereby reducing duplication of effort.

#### Area 4 – Standardisation, benchmarking and indexing of computational tools

The IDP research field is currently developing best practices for scientific (i/o) file formats, data analysis pipelines and benchmarking of scientific tools. These steps are being taken to raise the quality and accessibility of software developed by the community to produce more accurate, faster, more stable and user-friendly software implementation. The development and adoption of the storage and dissemination formats in *Priority Area 1* will help standardise the (i/o) file formats of IDP tools. However, several use cases will not be covered by these formats (e.g. standardised i/o formats for residue-specific IDR scoring) and additional effort will be required to formalise the output formats in these areas. Containerised software, such as those available via [BioContainers](#), or package managers, such as [BioConda](#), are rarely used in IDP research. However, such advances would make IDP tools more accessible to the wider community and would simplify their benchmarking.

The field of IDP research could also benefit from the development of reusable experimental workflows such as those implemented with the [Common Workflow Language](#) for commonly used IDP analysis pipelines built on top of [ELIXIR CDRs and Deposition Databases](#). These workflows could then be managed using the [Galaxy workflow manager platform](#). Benchmarking remains an issue for the community as the lack of common benchmarking datasets has hampered the systematic assessment of IDP tools. As a result, many publications have claimed superior performance for biased datasets, furthering the need for standardisation. Due to the availability of high-quality manually curated IDP data from DisProt, the community can provide gold standard blind datasets to run periodic benchmarking assessments of IDR prediction tools. This approach has recently been successfully applied by the [Critical Assessment of Intrinsic Disorder \(CAID\) initiative](#). A similar platform for the comparison of methods predicting functional modules within IDPs or aligning homologous IDRs would also benefit the community and drive advances in these developing fields.

The development of Open Source software to benchmark IDP analysis methods across a wide range of performance metrics will simplify and standardise the assessment of IDP tools.

[OpenEbench](#) can be used as an information hub to distribute reference datasets, to run comparative assessments and to publish benchmarking results. From a technical point of view, the adoption of BioContainers and/or Galaxy, and the standardisation of input and output formats would streamline the assessment process. Benchmarking results covering both scientific and technical aspects of the available IDP analysis tools can be hosted at OpenEBench to simplify method selection by users. Finally, the addition of these computational tools for the analysis of IDPs to the [bio.tools registry](#) would improve the visibility of these tools.

#### Area 5 – Development of a centralised knowledge-base for IDP data

Computational IDP researchers based in Europe develop many of the tools and resources that underlie the global IDP e-infrastructure. However, these assets are currently spread over numerous institutes and universities across Europe. The development of an umbrella resource, [DisProtCentral](#) (founded June 2017), consolidating the European IDP resources and tools through a single portal will improve the accessibility of these resources for the wider biological community. The DisProtCentral consortium will provide a central hub to access high-quality curation, annotation and predictions of structural and functional information on IDPs, in addition to providing stable identifiers/URIs to describe regions of disorder within specific proteins, enabling cross-referencing and linking between resources. The consortium will include all the stakeholders in the IDP field and provide a centralised repository for the protein disorder-related databases and tools. This will future-proof these key IDP resources against issues arising from the loss of funding or change of group focus.

This initiative to produce such a centralised resource draws attention to the need for a fit-for-purpose interoperability tooling adaptor that translates across the metadata annotations of individual resources. This will allow pre-existing data from distinct resources to be integrated via normalisation functionalities such as ontology cross-referencing for data previously mapped to different ontologies/vocabularies, or ontology term mapping for those that have not been mapped to any standard terminologies. Ideally, newly generated data should be produced FAIR-at-source (Findable, Accessible, Interoperable, and Reusable) (Wilkinson *et al.*, 2016). DisProtCentral will play a vital role in encouraging IDP data generators to follow the standardised set of best practices: identifying recommended ontologies that are universally adopted by those data generators, defining a common schema markup strategy, and validating persistent identifiers for data to be deposited into the resource. A wider goal of the DisProtCentral resource is to provide a single point of contact to promote discourse and collaboration to ensure that the needs of the IDP data generators, users and consumers are all being met by the computational IDP researchers. An important aspect of this effort will be the development of training material for the wider biological communities describing the best practices for IDP analyses.

## Alignment with ELIXIR Activities

The community challenges identified are already well-aligned with on-going ELIXIR platform goals and activities. In particular:

**Data Platform:** The task proposed in Priority Area 3 “Integration with ELIXIR Core Data Resources” aligns with the goals of the Data Platform. The community is already in the process of integrating IDP annotation into the ELIXIR CDRs as part of two ELIXIR implementation studies, “*Implementation study for the integration of ELIXIR-IIB in ELIXIR Data Curation activities*” and “*Integration and standardisation of intrinsically disordered protein data*”. The recent integration of MobiDB into InterPro, PDBE and UniProt can be taken as a blueprint for the further integration of IDP information into ELIXIR CDRs. Furthermore, initial discussions have taken place to plan the integration of additional sources of IDP data. A separate task, which can benefit from ongoing Data Platform activities, is the development of an IDP curation framework as described in Priority Area 2 “Automated and community-driven curation”. This aligns with ELIXIR Data Platform Task 3 “Scalable curation” of the 2019–2023 ELIXIR Programme.

The distributed community curation for IDP resources will benefit from stronger interaction with Europe PMC, with the SciLite framework providing literature mining. This is especially important where data is scattered throughout the extensive corpus of biological literature and not properly indexed. Together with an automated curation triage step selecting relevant papers, this can boost the productivity of community curation while ensuring high-quality annotation of the IDP literature. Initial efforts have started with the design of a dedicated curation-support prototype, which has been used by DisProt curators since 2018. The service capitalises on the neXtA5 platform (Mottin *et al.*, 2017), designed by the Text Mining group of the Swiss Institute of Bioinformatics. The demonstrator (<http://candy.hesge.ch/disprotGUI/>) is able to rank articles based on a scoring function, which prioritises articles with a high density of IDP-related concepts. Thanks to the curation-support tool, it will be possible to obtain a high quality curated benchmark, with the aim to evolve the current system from a triage system to a more accurate binary classifier, being able to select not only relevant papers but also to highlight short passages of text in full-text papers likely to support the expert curation.

**Tools Platform:** The tasks proposed in Priority Area 4 “Standardisation, benchmarking and indexing of computational tools” fall under the objectives of the Tools Platform. A major goal of Priority Area 4 is to increase productivity by developing reusable experimental workflows for commonly used IDP analysis pipelines built on top of ELIXIR CDRs. The Tools Platform can advise the ELIXIR IDP Community on the development of such workflows using the Common Workflow Language in collaboration with the ELIXIR Interoperability platform. The IDP analysis tool benchmarking by the CAID initiative is complementary to the OpenEBench benchmarking platform.

OpenEBench represents an information hub for the distribution of reference datasets, the application of metrics and comparative assessments and the distribution of benchmarking results. CAID can also drive prototype development for software containers and reusable experimental workflows for the community. The outcomes could then be generalised for other applications. The bio.tools service registry can be a comprehensive census of the available software for IDP analysis and the BioContainers and Bioconda platforms can assist in the generation of software containers and their registry across different technologies to facilitate access and use of software by the IDP communities and beyond.

**Interoperability Platform:** Several of the identified priorities will benefit from the input of the Interoperability Platform. The main interaction with the platform will be to develop FAIR-compliant standards and guidelines for reporting, data exchange and retrieval of data on IDP structure and function (Wilkinson *et al.*, 2016). The recent development of a HUPO PSI-ID workgroup and a draft MIADE standard are key steps towards this goal. IDP resources themselves will benefit from becoming FAIR and adhering and contributing to the further development of standards, controlled vocabularies and ontologies. These advances are of paramount importance to improve the dissemination of IDP data. FAIR-compliant data will also aid in the FAIRification of analysis tools where reusable computational workflows can make programmatic calls to retrieve reusable data from an integrated centralised source. The Interoperability Platform service framework has already defined the various key activities to support the “FAIRification” of data which can be applied to IDP data. The effort to make IDP data FAIR can exploit the various key services that are offered by the Interoperability Platform such as those identified as recommended practices through the **ELIXIR Recommended Interoperability Resources (RIRs)**, and the platform mission-critical initiatives (i.e., Bioschemas, and Common Workflow Language). As this is an emerging community with new data resources in development, deploying these recommended practices in new IDP databases will highlight the IDP Community as a use case example in the mission to make data *FAIR-at-source*. This work can be supported by the Interoperability Platform under their remit to provide support for interoperability for ELIXIR Communities.

**Compute Platform:** Work being carried out in the Compute Platform can be leveraged in two main ways. Community curation requires authentication for a wide range of participants, which can benefit from the federated ELIXIR AAI approach for curator login. This could also facilitate the attribution of credit to curators. A second element would be to use the ELIXIR distributed computing infrastructure including identity and access management, data integration and container deployment across a range of appropriate compute endpoints as required, both for updating large-scale databases and to run the CAID experiment. A range of new Compute Platform activities are underway to enable an integrated approach to the deployment of relevant workflows and containers. The IDP Community can take advantage of

these activities to deploy IDP-related workflows and containers.

**Training Platform:** The IDP Community is focussed on the development of the next generation of IDP researchers. Extensive training has been funded and carried out by the COST Action NGP-NET and additional training schools are planned as part of the MSCA-RISE-funded IDPfun project. The Italian ELIXIR node, which has been actively providing training, will organise a yearly training course on IDP resources. The input of the ELIXIR Training Platform will be indispensable to facilitate the development of a core set of training materials, its dissemination and inclusion in ELIXIR training courses. These materials will comprise introductory and advanced lessons on databases access, *in silico* analysis of IDPs (e.g. IDP prediction, IDP docking, MD simulations and IDP-specific primary sequence analysis techniques), and computational processing of experimental data (e.g. NMR spectra, phage display). There will also be a requirement for the development of training material related to IDP interoperability to train experimentalists, software developers, database developers and curators in topics such as IDP standards, IDP controlled vocabularies and Common Workflow Language (CWL) for IDPs. This aligns with the Interoperability Platform's training and outreach task in the 2019–2023 ELIXIR Programme to collaborate with Training Platform members to provide Interoperability training. In collaboration with the Training Platform, the IDP Community will be able to adopt and implement a wide range of best practices and guidelines available through the ELIXIR Training Toolkit, which aims to provide a comprehensive reference resource for developing training capacity, rolling out new training programs, as well as expanding existing ones. All materials and activities developed for IDP training will be shared through [TeSS](#) for further dissemination. Additionally, train-the-trainer activities will be planned to increase the size of the IDP trainers pool and build additional training capacity.

## Alignment with other communities

### Links to ELIXIR Communities

**Structural Bioinformatics (3DBioInfo):** ELIXIR 3DBioInfo is a community of structural bioinformaticians with the remit of continuing the development of the e-infrastructure for the storage, visualisation, analysis, annotation, and prediction of the structure of biological macromolecules and complexes. The probabilistic approach to protein structure employed by the IDP researchers is highly complementary to the work of the ELIXIR 3DBioInfo community and together, a complete structural description of proteins can be achieved. With the two communities addressing different challenges, there are numerous opportunities for synergistic connections between them, particularly in the development of structural ontologies and standards, structure boundary definition for structural studies, methods to predict IDR-globular domain interactions and whole protein structure prediction tools.

**Proteomics:** The ELIXIR Proteomics Community aspires to improve the research on proteoforms including protein

forms caused by post-translational modifications (PTMs) and sequence variants. To this end, proteoform-centric annotations of proteomes are needed. This includes information on the co-occurrences of IDRs with PTMs and non-constitutive exons, and if possible, the functional outcomes of these conditional changes to the protein. Close collaboration is desirable between the communities regarding data interoperability to promote the use of IDP and proteomics data in data analysis workflows. IDP research also applies proteomics-related technologies such as ion-mobility mass spectrometry and cross-linking mass spectrometry. In these cases, close collaboration with the proteomics community within the HUPO-PSI Mass Spectrometry workgroup will support the standardisation efforts of the IDP community.

**Rare Diseases:** The overarching goal of the ELIXIR Rare Diseases Community is to create a sustainable, reusable, and interoperable infrastructure that will enable researchers to discover, access, and analyse rare disease data. A key aspect of the analysis of rare diseases data is the elucidation of the mechanism(s) by which genetic change(s) result in a diseased state. While the interpretation of the effect of disease mutations altering globular domains is well established, IDPs represent challenges in this context. Interaction and collaboration with the ELIXIR IDP Community will provide invaluable information to aid in the understanding of disease-causing variations, truncations, and/or chimeric oncogene(s) related to IDRs. The development, and dissemination of standards for IDR structural and functional data as proposed in priority area 1 are important aspects for the collaboration between the ELIXIR IDP and Rare Diseases Communities.

**Galaxy:** Galaxy is a web-based e-infrastructure for computational biomedical research. It allows users with minimal computational proficiency to run and share data analysis workflows. This promotes reproducibility and simplifies sharing of data and results. Currently, there is limited use of Galaxy by the ELIXIR IDP community. However, one of the priority areas of the roadmap, “*Standardisation, benchmarking and indexing of computational tools*”, is the development of reusable experimental workflows for commonly used IDP analysis pipelines built on top of ELIXIR CDRs. A collaboration with the ELIXIR Galaxy Community would be highly beneficial for such an endeavour, particularly in the development of IDP pipelines using the ELIXIR scientific workflow platform.

### Links to non-ELIXIR communities

**Instruct-ERIC:** Instruct-ERIC is a European research infrastructure for structural biology that makes high-end technologies and methods available to European researchers. Similarly to the proposed ELIXIR 3DBioInfo Community, the interdependency of the research performed by Instruct-ERIC and the IDP researchers, covering the structured and unstructured parts of the proteome, allows for extensive synergies. Instruct-ERIC coordinates access to facilities which permit the structural analysis of IDPs including several centres specialised in NMR and biophysical techniques. Furthermore, Instruct-ERIC has developed a set of high-quality training courses in structural



biology. Given the complementary teaching focus of the ELIXIR IDP Community and the Instruct-ERIC initiative, future training schools with contributions from both sources would be of great benefit to any participant.

**The Dark Proteome Initiative:** The Dark Proteome Initiative is a US consortium of experimentalists with the goal of fostering collaborations amongst IDP researchers and lobbying for funding to address open and biologically important questions in the IDP field. The Dark Proteome Initiative is a readymade community of world-class generators of IDP data utilising a wide range of experimental approaches allowing the ELIXIR IDP Community to access a single entity for guidance on the needs of experimental data generators. The Dark Proteome Initiative will also be able to provide invaluable advice on the detailed definitions and requirements for the development of ontologies and standards.

**PLUMED consortium:** The PLUMED consortium has been recently established as an open community including PLUMED developers, contributors and users to help to

establish reproducibility, open access data, and harmonisation of the protocols for molecular dynamics simulations, free energy calculations and other simulations that can be run with the PLUMED software (PLUMED consortium, 2019).

## Conclusions

Recent years have seen a rapid growth of interest in IDPs. This has coincided with significant advances in the *in vivo*, *in vitro* and *in silico* methods to study the structure and function of IDPs. Unfortunately, the basic requirements for the organisation and dissemination of the tools and data produced by the field have not advanced in step with these developments. The field is on the cusp of an era where the high-throughput characterisation of the extensive intrinsically disordered regions (IDRs) of proteins is possible. This roadmap will provide the foundation that supports this data explosion and provide a solid platform for the future biological research of IDPs.

## Data availability

No data are associated with this article.

## References

- Babu MM, Kriwacki RW, Pappu RV: **Structural biology. Versatility from protein disorder.** *Science*. 2012; **337**(6101): 1460–1461.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Baggett DW, Nath A: **The Rational Discovery of a Tau Aggregation Inhibitor.** *Biochemistry*. 2018; **57**(42): 6099–6107.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bah A, Forman-Kay JD: **Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications.** *J Biol Chem*. 2016; **291**(13): 6696–6705.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Beltrao P, Albanese V, Kenner LR, et al.: **Systematic functional prioritization of protein posttranslational modifications.** *Cell*. 2012; **150**(2): 413–425.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bernadó P, Mylonas E, Petoukhov MV, et al.: **Structural characterization of flexible proteins using small-angle X-ray scattering.** *J Am Chem Soc*. 2007; **129**(17): 5656–5664.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Best RB: **Computational and theoretical advances in studies of intrinsically disordered proteins.** *Curr Opin Struct Biol*. 2017; **42**: 147–154.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Blikstad C, Ivarsson Y: **High-throughput methods for identification of protein-protein interactions involving short linear motifs.** *Cell Commun Signal*. 2015; **13**: 38.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bonetti D, Troilo F, Brunori M, et al.: **How Robust Is the Mechanism of Folding-Up-On-Binding for an Intrinsically Disordered Protein?** *Biophys J*. 2018; **114**(8): 1889–1894.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Borgia A, Borgia MB, Bugge K, et al.: **Extreme disorder in an ultrahigh-affinity protein complex.** *Nature*. 2018; **555**(7694): 61–66.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Britan A, Cusin I, Hinard V, et al.: **Accelerating annotation of articles via automated approaches: evaluation of the neXtA<sub>2</sub> curation-support tool by neXtProt.** *Database (Oxford)*. 2018; **2018**.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Buljan M, Chalancon G, Eustermann S, et al.: **Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks.** *Mol Cell*. 2012; **46**(6): 871–883.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Camilloni C, De Simone A, Vranken WF, et al.: **Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts.** *Biochemistry*. 2012; **51**(11): 2224–2231.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chong SH, Chatterjee P, Ham S: **Computer Simulations of Intrinsically Disordered Proteins.** *Annu Rev Phys Chem*. 2017; **68**: 117–134.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chouard T: **Structural biology: Breaking the protein rules.** *Nature*. 2011; **471**(7337): 151–153.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cilia E, Pancsa R, Tompa P, et al.: **The DynaMine webserver: predicting protein dynamics from sequence.** *Nucleic Acids Res*. 2014; **42**(Web Server issue): W264–70.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Corbi-Verge C, Kim PM: **Motif mediated protein-protein interactions as drug targets.** *Cell Commun Signal*. 2016; **14**: 8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Csizmok V, Forman-Kay JD: **Complex regulatory mechanisms mediated by the interplay of multiple post-translational modifications.** *Curr Opin Struct Biol*. 2018; **48**: 58–67.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Davey NE: **The functional importance of structure in unstructured protein regions.** *Curr Opin Struct Biol*. 2019; **56**: 155–163.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Davey NE, Cyert MS, Moses AM: **Short linear motifs - ex nihilo evolution of protein regulation.** *Cell Commun Signal*. 2015; **13**: 43.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Davey NE, Morgan DO: **Building a Regulatory Network with Short Linear Sequence Motifs: Lessons from the Degrons of the Anaphase-Promoting Complex.** *Mol Cell*. 2016; **64**(1): 12–23.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Davey NE, Seo MH, Yadav VK, et al.: **Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome.** *FEBS J*. 2017; **284**(3): 485–498.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Davey NE, Travé G, Gibson TJ: **How viruses hijack cell regulation.** *Trends Biochem Sci*. 2011; **36**(3): 159–169.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dosztányi Z, Csizmok V, Tompa P, et al.: **The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins.** *J Mol Biol*. 2005; **347**(4): 827–839.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dosztányi Z, Mészáros B, Simon I: **ANCHOR: web server for predicting protein binding regions in disordered proteins.** *Bioinformatics*. 2009; **25**(20): 2745–2746.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dunker AK, Silman I, Uversky VN, et al.: **Function and structure of inherently disordered proteins.** *Curr Opin Struct Biol*. 2008; **18**(6): 756–64.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Durinx C, McEntyre J, Appel R, et al.: **Identifying ELIXIR core data resources [version 2; peer review: 2 approved].** *F1000Res*. 2016; **5**: pii: ELIXIR-2422.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dyson HJ, Wright PE: **How do intrinsically disordered viral proteins hijack the**



- cell? *Biochemistry*. 2018; **57**(28): 4045–4046.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dyson HJ, Wright PE: **Intrinsically unstructured proteins and their functions.** *Nat Rev Mol Cell Biol*. 2005; **6**(3): 197–208.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Edwards RJ, Palopoli N: **Computational prediction of short linear motifs from protein sequences.** *Methods Mol Biol*. 2015; **1268**: 89–141.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Felli IC, Pierattelli R: **Intrinsically Disordered Proteins Studied by NMR Spectroscopy.** Springer. 2015.  
[Publisher Full Text](#)
- Forman-Kay JD, Mittag T: **From sequence and forces to structure, function, and evolution of intrinsically disordered proteins.** *Structure*. 2013; **21**(9): 1492–1499.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Franko D, Petoukhov MV, Konarev PV, *et al.*: **ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions.** *J Appl Crystallogr*. 2017; **50**(Pt 4): 1212–1225.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fuertes G, Banterle N, Ruff KM, *et al.*: **Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements.** *Proc Natl Acad Sci U S A*. 2017; **114**(31): E6342–E6351.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fuxreiter M: **Fold or not to fold upon binding - does it really matter?** *Curr Opin Struct Biol*. 2018; **54**: 19–25.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gibson TJ, Dinkel H, Van Roey K, *et al.*: **Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad.** *Cell Commun Signal*. 2015; **13**: 42.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gouw M, Michael S, Sámano-Sánchez H, *et al.*: **The eukaryotic linear motif resource - 2018 update.** *Nucleic Acids Res*. 2018; **46**(D1): D428–D434.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Guharoy M, Bhowmick P, Sallam M, *et al.*: **Tripartite degrons confer diversity and specificity on regulated protein degradation in the ubiquitin-proteasome system.** *Nat Commun*. 2016; **7**: 10239.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hess B, Kutzner C, van der Spoel D, *et al.*: **GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.** *J Chem Theory Comput*. 2008; **4**(3): 435–47.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Holehouse AS, Pappu RV: **Collapse Transitions of Proteins and the Interplay Among Backbone, Sidechain, and Solvent Interactions.** *Annu Rev Biophys*. 2018; **47**: 19–39.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Holmstrom ED, Holla A, Zheng W, *et al.*: **Accurate Transfer Efficiencies, Distance Distributions, and Ensembles of Unfolded and Intrinsically Disordered Proteins From Single-Molecule FRET.** *Methods Enzymol*. 2018; **611**: 287–325.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Huang J, MacKerell AD Jr: **Force field development and simulations of intrinsically disordered proteins.** *Curr Opin Struct Biol*. 2018; **48**: 40–48.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Iakoucheva LM, Radivojac P, Brown CJ, *et al.*: **The importance of intrinsic disorder for protein phosphorylation.** *Nucleic Acids Res*. 2004; **32**(3): 1037–1049.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Iešmantavicius V, Dogan J, Jemth P, *et al.*: **Helical propensity in an intrinsically disordered protein accelerates ligand binding.** *Angew Chem Int Ed Engl*. 2014; **53**(6): 1548–1551.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ivarsson Y, Jemth P: **Affinity and specificity of motif-based protein-protein interactions.** *Curr Opin Struct Biol*. 2019; **54**: 26–33.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kriwacki RW, Hengst L, Tennant L, *et al.*: **Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity.** *Proc Natl Acad Sci U S A*. 1996; **93**(21): 11504–11509.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kruse T, Biedenkopf N, Hertz EPT, *et al.*: **The Ebola Virus Nucleoprotein Recruits the Host PP2A-B56 Phosphatase to Activate Transcriptional Support Activity of VP30.** *Mol Cell*. 2019; **69**(1): 136–145.e6.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Krystkowiak I, Davey NE: **SLIMSearch: a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions.** *Nucleic Acids Res*. 2017; **45**(W1): W464–W469.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Light S, Sagit R, Satchenkova O, *et al.*: **Protein expansion is primarily due to indels in intrinsically disordered regions.** *Mol Biol Evol*. 2013; **30**(12): 2645–2653.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Martin EW, Mittag T: **Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions.** *Biochemistry*. 2018; **57**(17): 2478–2487.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mészáros B, Kumar M, Gibson TJ, *et al.*: **Degrins in cancer.** *Sci Signal*. 2017; **10**(470): pii: eaak9982.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Milles S, Salvi N, Blackledge M, *et al.*: **Characterization of intrinsically disordered proteins and their dynamic complexes: From *in vitro* to cell-like environments.** *Prog Nucl Magn Reson Spectrosc*. 2018; **109**: 79–100.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mir S, Alhroub Y, Anyango S, *et al.*: **PDBe: towards reusable data delivery infrastructure at protein data bank in Europe.** *Nucleic Acids Res*. 2018; **46**(D1): D486–D492.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mitreá DM, Kriwacki RW: **Phase separation in biology: functional organization of a higher order.** *Cell Commun Signal*. 2016; **14**: 1.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mottin L, Pasche E, Gobeill J, *et al.*: **Triage by ranking to support the curation of protein interaction.** *Database (Oxford)*. 2017; 2017.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mottin L, Pasche E, Gobeill J, *et al.*: **Triage by ranking to support the curation of protein interactions.** *Database (Oxford)*. 2017; 2017.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Necci M, Piovesan D, Dosztányi Z, *et al.*: **MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins.** *Bioinformatics*. 2017; **33**(9): 1402–1404.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Neduvá V, Russell RB: **DILIMOT: discovery of linear motifs in proteins.** *Nucleic Acids Res*. 2006; **34**(Web Server issue): W350–5.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nguyen HQ, Baxter BC, Brower K, *et al.*: **Programmable Microfluidic Synthesis of Over One Thousand Uniquely Identifiable Spectral Codes.** *Adv Opt Mater*. 2017; **5**(3): pii: 1600548.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nielsen JT, Mulder FAA: **There is Diversity in Disorder-"In all Chaos there is a Cosmos, in all Disorder a Secret Order".** *Front Mol Biosci*. 2016; **3**: 4.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nodet G, Salmon L, Ozenne V, *et al.*: **Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings.** *J Am Chem Soc*. 2009; **131**(49): 17908–17918.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Oates ME, Romero P, Ishida T, *et al.*: **D<sup>2</sup>P: database of disordered protein predictions.** *Nucleic Acids Res*. 2013; **41**(Database issue): D508–16.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Olsen JG, Teilum K, Kragelund BB: **Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness.** *Cell Mol Life Sci*. 2017; **74**(17): 3175–3183.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Orchard S, Ammari M, Aranda B, *et al.*: **The MintAct project-IntAct as a common curation platform for 11 molecular interaction databases.** *Nucleic Acids Res*. 2014; **42**(Database issue): D358–63.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Orchard S, Kerrien S, Abbani S, *et al.*: **Protein interaction data curation: the International Molecular Exchange (IMEx) consortium.** *Nat Methods*. 2012; **9**(4): 345–350.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ozenne V, Bauer F, Salmon L, *et al.*: **Flexible-mecano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables.** *Bioinformatics*. 2012; **28**(11): 1463–1470.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Pancsa R, Tompa P: **Structural disorder in eukaryotes.** *PLoS One*. 2012; **7**(4): e34687.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Peng Y, Cao S, Kiselar J, *et al.*: **A Metastable Contact and Structural Disorder in the Estrogen Receptor Transactivation Domain.** *Structure*. 2019; **27**(2): 229–240.e4.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Piovesan D, Tabaro F, Mičetić I, *et al.*: **DisProt 7.0: a major update of the database of disordered proteins.** *Nucleic Acids Res*. 2017; **45**(D1): D219–D227.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Piovesan D, Tabaro F, Paladin L, *et al.*: **MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins.** *Nucleic Acids Res*. 2018; **46**(D1): D471–D476.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Plitzko JM, Schuler B, Selenko P: **Structural Biology outside the box-inside the cell.** *Curr Opin Struct Biol*. 2017; **46**: 110–121.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Prilusky J, Felder CE, Zeev-Ben-Mordehai T, *et al.*: **FoldIndex@: a simple tool to predict whether a given protein sequence is intrinsically unfolded.** *Bioinformatics*. 2005; **21**(16): 3435–3438.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Raveh B, London N, Zimmerman L, *et al.*: **Rosetta FlexPepDock *ab-initio*: simultaneous folding, docking and refinement of peptides onto their receptors.** *PLoS One*. 2011; **6**(4): e18934.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Santofimia-Castaño P, Xia Y, Lan W, *et al.*: **Ligand-based design identifies a potent NUPR1 inhibitor exerting anticancer activity via necroptosis.** *J Clin Invest*. 2019; **129**(6): 2500–2513.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

- Schad E, Fichó E, Pancsa R, *et al.*: **DIBS: a repository of disordered binding sites mediating interactions with ordered proteins.** *Bioinformatics*. 2018; **34**(3): 535–537.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schuler B, Soranno A, Hofmann H, *et al.*: **Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins.** *Annu Rev Biophys*. 2016; **45**: 207–231.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Scott JD, Pawson T: **Cell signaling in space and time: where proteins come together and when they're apart.** *Science*. 2009; **326**(5957): 1220–1224.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shigemitsu Y, Hiroaki H: **Common molecular pathogenesis of disease-related intrinsically disordered proteins revealed by NMR analysis.** *J Biochem*. 2018; **163**(1): 11–18.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Shoemaker BA, Portman JJ, Wolynes PG: **Speeding molecular recognition by using the folding funnel: the fly-casting mechanism.** *Proc Natl Acad Sci U S A*. 2000; **97**(16): 8868–8873.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sickmeier M, Hamilton JA, LeGall T, *et al.*: **DisProt: the Database of Disordered Proteins.** *Nucleic Acids Res*. 2007; **35**(Database issue): D786–93.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sivade Dumousseau M, Alonso-López D, Ammari M, *et al.*: **Encompassing new use cases - level 3.0 of the HUPO-PSI format for molecular interactions.** *BMC Bioinformatics*. 2018; **19**(1): 134.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Stanley N, Esteban-Martín S, De Fabritius G: **Progress in studying intrinsically disordered proteins with atomistic simulations.** *Prog Biophys Mol Biol*. 2015; **119**(1): 47–52.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tamiola K, Mulder FA: **Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins.** *Biochem Soc Trans*. 2012; **40**(5): 1014–1020.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tolchard J, Walpole SJ, Miles AJ, *et al.*: **The intrinsically disordered Tarp protein from chlamydia binds actin with a partially preformed helix.** *Sci Rep*. 2018; **8**(1): 1960.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tomba P: **The interplay between structure and function in intrinsically unstructured proteins.** *FEBS Lett*. 2005; **579**(15): 3346–3354.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tomba P: **Unstructural biology coming of age.** *Curr Opin Struct Biol*. 2011; **21**(3): 419–425.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tomba P, Davey NE, Gibson TJ, *et al.*: **A million peptide motifs for the molecular biologist.** *Mol Cell*. 2014; **55**(2): 161–169.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Trabuco LG, Lise S, Petsalaki E, *et al.*: **PepSite: prediction of peptide-binding sites from protein surfaces.** *Nucleic Acids Res*. 2012; **40**(Web Server issue): W423–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tribello GA, Bonomi M, Branduardi D, *et al.*: **PLUMED2: New feathers for an old bird.** *Comp Phys Comm*. 2014; **185**(2): 604–613.  
[Publisher Full Text](#)
- Ulrich EL, Akutsu H, Doreleijers JF, *et al.*: **BioMagResBank.** *Nucleic Acids Res*. 2008; **36**(Database issue): D402–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- UniProt Consortium: **UniProt: a worldwide hub of protein knowledge.** *Nucleic Acids Res*. 2019; **47**(D1): D506–D515.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Uversky VN, Oldfield CJ, Dunker AK: **Intrinsically disordered proteins in human diseases: introducing the D<sup>2</sup> concept.** *Annu Rev Biophys*. 2008; **37**: 215–246.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Valentini E, Kikhney AG, Previtali G, *et al.*: **SASBDB, a repository for biological small-angle scattering data.** *Nucleic Acids Res*. 2015; **43**(Database issue): D357–63.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vallat B, Webb B, Westbrook JD, *et al.*: **Development of a Prototype System for Archiving Integrative/Hybrid Structure Models of Biological Macromolecules.** *Structure*. 2018; **26**(6): 894–904.e2.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- van der Lee R, Buljan M, Lang B, *et al.*: **Classification of intrinsically disordered regions and proteins.** *Chem Rev*. 2014; **114**(13): 6589–6631.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Van Roey K, Dinkel H, Weatheritt RJ, *et al.*: **The switches.ELM resource: a compendium of conditional regulatory interaction interfaces.** *Sci Signal*. 2013; **6**(269): rs7.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Van Roey K, Gibson TJ, Davey NE: **Motif switches: decision-making in cell regulation.** *Curr Opin Struct Biol*. 2012; **22**(3): 378–385.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Van Roey K, Uyar B, Weatheritt RJ, *et al.*: **Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation.** *Chem Rev*. 2014; **114**(13): 6733–6778.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Varadi M, Kosol S, Lebrun P, *et al.*: **pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins.** *Nucleic Acids Res*. 2014; **42**(Database issue): D326–35.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Via A, Uyar B, Brun C, *et al.*: **How pathogens use linear motifs to perturb host cell networks.** *Trends Biochem Sci*. 2015; **40**(1): 36–48.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Volkmer R: **Synthesis and application of peptide arrays: quo vadis SPOT technology.** *ChemBiochem*. 2009; **10**(9): 1431–1442.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Walsh I, Martin AJ, Di Domenico T, *et al.*: **ESpritz: accurate and fast prediction of protein disorder.** *Bioinformatics*. 2012; **28**(4): 503–509.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Weatheritt RJ, Davey NE, Gibson TJ: **Linear motifs confer functional diversity onto splice variants.** *Nucleic Acids Res*. 2012; **40**(15): 7123–7131.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Weatheritt RJ, Gibson TJ: **Linear motifs: lost in (pre)translation.** *Trends Biochem Sci*. 2012; **37**(8): 333–341.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Whitmore L, Miles AJ, Mavridis L, *et al.*: **PCDDb: new developments at the Protein Circular Dichroism Data Bank.** *Nucleic Acids Res*. 2017; **45**(D1): D303–D307.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci Data*. 2016; **3**: 160018.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wright PE, Dyson HJ: **Intrinsically disordered proteins in cellular signalling and regulation.** *Nat Rev Mol Cell Biol*. 2015; **16**(1): 18–29.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Xue B, Dunker AK, Uversky VN: **Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life.** *J Biomol Struct Dyn*. 2012; **30**(2): 137–149.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Xue B, Mizianty MJ, Kurgan L, *et al.*: **Protein intrinsic disorder as a flexible armor and a weapon of HIV-1.** *Cell Mol Life Sci*. 2012; **69**(8): 1211–1259.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Yu C, Niu X, Jin F, *et al.*: **Structure-based Inhibitor Design for the Intrinsically Disordered Protein c-Myc.** *Sci Rep*. 2016; **6**: 22298.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 25 November 2019

<https://doi.org/10.5256/f1000research.22111.r55395>

© 2019 Ugur Sezerman O. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Osman Ugur Sezerman**

Acibadem University, Istanbul, Turkey

The paper describes importance of intrinsically disordered proteins with thorough set of examples with good references. It also explains computational and experimental challenges involved in IDP research.

A good summary of available tools and platforms are also provided. Key challenges in IDP research are identified and key activities to be carried out to tackle these challenges are provided. ELIXIR community has many of the researchers working in the field within the consortium. All the needed activities align well with ELIXIR activities as well. So it is natural that ELIXIR plays a centralized role in these activities. The paper also identified key players around the world that they can collaborate with establishing a task force to tackle major challenges IDP research needs to tackle.

Only issue I have with the paper is the lack of focus on reproducibility issues. That is a major problem life science research is facing today. most of the experimental results are not reproducible even computational models are not reproducible resulting in waste of billions of euros spent in funding such research activities.

They should have a special focus group within their consortium to ensure reproducibility of experimental results within IDP community (define key information needed to reproduce results, endorse use of e notebooks and use of protocols established withing the network, etc.). They should also do the same for computational models that are developed to tackle several problems related to IDP. A section added on these issues would strengthen the paper. Also if they include other initiatives focusing on reproducibility that would strengthen their task force.

**Is the topic of the opinion article discussed accurately in the context of the current literature?**

Yes

**Are all factual statements correct and adequately supported by citations?**

Yes

**Are arguments sufficiently supported by evidence from the published literature?**

Yes

**Are the conclusions drawn balanced and justified on the basis of the presented arguments?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** bioinformatics, system biology, structural bioinformatics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 21 November 2019

<https://doi.org/10.5256/f1000research.22111.r55886>

© 2019 Forman-Kay J et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Iva Pritisanac** 

<sup>1</sup> Molecular Medicine, Hospital for Sick Children, Toronto, ON, Canada

<sup>2</sup> Cell & Systems Biology, University of Toronto, Toronto, ON, Canada

**Julie Forman-Kay**

<sup>1</sup> Molecular Medicine, Hospital for Sick Children, Toronto, ON, Canada

<sup>2</sup> Biochemistry, University of Toronto, ON, Canada

There has been a significant contribution from Europe to the field of intrinsically disordered protein (IDP) research, including development of computational tools. This white paper recognizes that science and methodological development, and reports on a meeting focused on organizing the software and the software developers supporting the IDP community in Europe. While it is extremely important to provide infrastructure support for the IDP community and efforts to increase communication and usage of tools are essential, it is not clear that the significant effort to put these resources under a single umbrella and unify data formats is of value beyond that of encouraging independent development with opportunities to interact. This white paper provides a useful report on the meeting that will be of value to the IDP computational community. However, it could be improved to provide greater clarity on the proposal and some of the suggested priorities could be re-evaluated.

#### Specific comments:

1. The white paper itself is challenging to read. It is unclear who the target audience of the article is? Is it the IDP/IDR community or the ELIXIR community, or both? In the present form, the article might be confusing to both communities, as it provides a lot (at times too much) detail for some aspects, resources, and individual tools, while omitting crucial information for others. For instance, it is unclear what from the current resources would be relocated into a new ELIXIR infrastructure and what would continue to exist independently and simply being "pointed at" from a central resource? It is also unclear exactly where central data repositories and central software platforms would reside. Many presently available resources are outlined and suggested at different points, making it unclear how everything would fit together. In addition, it is unclear if there will be a central

resource(s), such as DisProtCentral, and if so would this resource(s) ultimately have independent developers, next to the participating research groups.

2. Some of the software and the databases that will be the key components are listed in Figure 2. However, not all of the outlined resources are presently functional or consistently maintained. It is therefore unclear if the initiative aims at reviving some of the resources in the future, and if so, whether this be done by the individual research groups. For the present time, it would be important to include a table with links to the websites/repositories of the operational resources, and a note on the current status of all other, e.g. "not actively developed", "under development", "under restructuring", etc.
3. It would be very helpful to include schematics to help understanding of the proposed initiative and improve the flow of the article. For instance, it would be good to use flow charts to outline the hierarchy of priorities to help understand how presently available resources fit into different priority areas and what will be the first tasks undertaken by different work groups. The priorities relating to the Area 1 ("Standards and exchange formats for IDP data") would be especially important to clarify in that sense, so that the community can be informed where to look for updates on the matter and thus help aid these efforts. Finally, the history of the interactions between the IDP community and ELIXIR and the subsequent integration into the ELIXIR core could be represented with a timeline and be in a separate subsection of the article, as this presently interrupts the flow of the article.
4. It would be important to define the strategy that the community wants to undertake towards integration of resources. For instance this could be a) integrated, dedicated databases with API access that are continuously maintained for data storage; and b) open source software with modular architecture that will enable integration into, e.g., BioConda for software development. At present, the general strategy is not clear or explicit, making the article quite confusing to follow.
5. It would be highly valuable to use this article and this platform to promote open source software development and established platforms for this purpose, such as GitHub. GitHub has proven effective in countless large-scale projects and it connects thousands of developers. The comparatively very small community of IDP computational researchers would benefit by developing, integrating, and sharing their resources through GitHub. In addition, such a platform would ensure inclusiveness.  
Related to this, it is not clear how the proposed centralization will solve the major issue of long term support for software, including funding for people to continue to maintain and develop the tools. A good way to assure a project will continue is for the project source/algorithm be fully open, in a repository such as GitHub, to enable use of modular code within other projects.
6. It is stated that DisProtCentral would be used for consolidating the European IDP resources and tools. However, it has not been mentioned what is the plan (if any) for integration of non-European resources and tools? Would there be some way for developers from outside of Europe to qualify to submit or include their software and resources in DisProtCentral?
7. There is a lot of emphasis on 'interoperability', and adaptors that could be used to this end. However, apart from the simplest workflows, e.g. annotation of predicted disordered regions on a protein sequence, it is likely that the demands for the workflows will be rapidly changing in the future. Spending effort on defining formats for data is not likely to be productive since by the time these standards can be applied they are could already be outdated. Technically speaking, it is not



a problem if data are used in different formats, as long as those formats are easy to parse, since writing a parser is trivial. Therefore, the emphasis should be shifted to the modular software development that would enable developers to 'mix and match' various routines and dynamically create their own workflows adapt to the task at hand. The proposal for integration into bio.tools or BioConda in this context is a very good one.

8. There are two tools that are actively maintained to current standards, GROMACS and PLUMED2, and they are not IDP specific and they run well regardless of ELIXIR. These points should be noted. Capabilities to download the software, see the development flow and interact with the project directly are available for these tools, but it is not clear whether these principles are a general priority.

**Minor points:**

1. A full name and abbreviation for ELIXIR should be given in the abstract.
2. International Molecular Exchange (IMEx) is mentioned in the article before the abbreviation is introduced (the abbrev. is introduced later).

**Is the topic of the opinion article discussed accurately in the context of the current literature?**

Yes

**Are all factual statements correct and adequately supported by citations?**

Yes

**Are arguments sufficiently supported by evidence from the published literature?**

Yes

**Are the conclusions drawn balanced and justified on the basis of the presented arguments?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** intrinsically disordered protein experimental biophysics, biochemistry and computational development

**We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**